SANDIA REPORT
SAND2007-5260
Unlimited Release
Printed August 2007

Final Report
PATTON Alliance Gazetteer Evaluation Project

Denise R. Bleakly

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2007-5260
Unlimited Release
Printed August 2007

Final Report
PATTON Alliance Gazetteer Evaluation Project

Denise R. Bleakly
Critical Information Systems Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0763

ABSTRACT

In 2005 the National Ground Intelligence Center (NGIC) proposed that the PATTON Alliance provide assistance in evaluating and obtaining the Integrated Gazetteer Database (IGDB), developed for the Naval Space Warfare Command Research group (SPAWAR) under Advance Research and Development Activity (ARDA) funds by MITRE Inc., fielded to the text-based search tool *GeoLocator*, currently in use by NGIC. We met with the developers of *GeoLocator* and identified their requirements for a better gazetteer. We then validated those requirements by reviewing the technical literature, meeting with other members of the intelligence community (IC), and talking with both the United States Geologic Survey (USGS) and the National Geospatial Intelligence Agency (NGA), the authoritative sources for official geographic name information. We thus identified 12 high-level requirements from users and the broader intelligence community. The IGDB satisfies many of these requirements. We identified gaps and proposed ways of closing these gaps.

Three important needs have not been addressed but are critical future needs for the broader intelligence community. These needs include standardization of gazetteer data, a web feature service for gazetteer information that is maintained by NGA and USGS but accessible to users, and a common forum that brings together IC stakeholders and federal agency representatives to provide input to these activities over the next several years. Establishing a robust gazetteer web feature service that is available to all IC users may go a long way toward resolving the gazetteer needs within the IC. Without a common forum to provide input and feedback, community adoption may take significantly longer than anticipated with resulting risks to the war fighter.

CONTACT INFORMATION

Prepared by the PATTON Alliance for the
National Ground Intelligence Center
Pathfinder Program Office

Denise Bleakly
Gazetteer Lead, The PATTON Alliance
Critical Information Systems Department
Sandia National Laboratories
P.O. Box 5800, MS 0763
Albuquerque, NM 87185-0763

Phone: (505) 284-2535
Fax: (505) 844-3322
Email: drbleak@sandia.gov

CONTENTS

TABLES

ACRONYMS AND ABBREVIATIONS

ADL	Alexandria Digital Library Project
ARDA	Advance Research and Development Activity
AQUAINT	Advanced Question Answering for Intelligence
BGN	Board on Geographic Names, same as USBGN
CIA	Central Intelligence Agency
COGNA	Council on Geographic Names Authority
DNC	Domestic Names Committee
FNC	Foreign Names Committee
GIS	Geographic Information System
GNIS	Geographic Names Information System, a domestic geographic place-name information system hosted by the United States Geological Survey
GNS	GeoNet Names Server, a foreign geographic place-name information system hosted by the National Geospatial Intelligence Agency
IC	intelligence community
IGDB	Integrated Gazetteer Database, a product produced by MITRE Corporation for SPAWAR
KML	Keyhole Mark-up Language, an XML-based language for modeling and storing such geographic features as points, lines (?), images, and polygons. KML is used by Google Earth to allow developers to extend the number of layers of information that can be displayed by Google Earth.
LoC	Library of Congress
NGA	National Geospatial Intelligence Agency
NGIC	National Ground Intelligence Center
PATTON	Preparing Analytic Tools and Technology for an Operational Network, an alliance and its programs that are hosted at Sandia National Laboratories
OGC	Open Geospatial Consortium
R&D	research and development
SPAWAR	Naval Space Warfare Command Research group
UN	United Nations
USBGN	United States Board on Geographic Names
USGS	United States Geologic Survey

1. INTRODUCTION

The Preparing Analytic Tools and Technology for an Operational Network (PATTON) Alliance substantially accelerates the transition of information technologies from government-funded research to the field. The central mission of the PATTON Alliance is to create a channel for deploying mature, federally funded technologies to government users, while avoiding the discontinuity and subsequent cost and loss of capability that often occurs between prototype development and system deployment. The PATTON Alliance provides feedback to its sponsors on the operational utility of candidate technologies that are tested and applied to operational problems. Furthermore, it actively seeks matches between mature technologies and interested stakeholders. In its efforts to find a home for new technologies in operational networks, the PATTON Alliance will enhance the technical capability of the supported communities. The motto of the Alliance could be: "We don't make the tools you use, we make the tools you use better."

One of the early projects of the PATTON Alliance is to evaluate the previously Advanced Research and Development Activity (ARDA) funded Integrated Gazetteer Database (IGDB) and identify the gaps between this research and development (R&D) effort for ARDA and the wider intelligence community (IC) needs for a gazetteer.

In 2005, the National Ground Intelligence Center (NGIC) proposed that the PATTON Alliance provide assistance in fielding the IGDB, developed for the Naval Space Warfare Command Research group (SPAWAR) under ARDA research funds by MITRE Inc., to the text-based search tool *GeoLocator*, currently in use by NGIC. In 2006, work began on this Gazetteer Evaluation project. We were given access to the government-funded research prototype so that we could evaluate it against the needs of the IC. We met with the developers of *GeoLocator* to identify their requirements for a better gazetteer. We then validated those requirements by reviewing the technical literature, meeting with other members of the IC, and talking with both the United States Geologic Survey (USGS) and the National Geospatial Intelligence Agency (NGA), the authoritative sources for official geographic name information.

2. GAZETTEER NEEDS

Within the IC, many efforts are underway to use natural language and text-based search tools to systematically analyze large volumes of text for pertinent information and then display the results using a map. A gazetteer is used to match place names or named geographic features with geographic coordinates for map display.

Originally we were tasked by NGIC to compare the gazetteer needs of the *GeoLocator* with the MITRE IGDB. However under the PATTON Alliance program guidelines, we were also asked to determine if any other systems used by the IC could benefit from the IDGB.

Our approach was a four-step process.

1. Initially, we obtained requirements for an improved gazetteer from NGIC.

2. We then talked with other IC members and obtained from them their requirements for gazetteers or gazetteer services.

3. We reviewed existing gazetteers, gazetteer services, and internet sources of gazetteer information.

4. We then prepared a preliminary gap analysis that compared the IGDB to the requirements gathered.

We did not interview the end-user analysts, but rather gazetteer users from an information systems perspective – developers, Geographic Information System (GIS) analysts, and information system administrators. This provided information from the systems perspective in that it provided us with data concerning how a gazetteer is integrated into a larger information processing system. Frank Linton, of MITRE Inc, did a study of end-user analysts and their use of the NGA GeoNames gazetteer. This paper (Linton, 2005) aided our work by providing us with detailed information on how gazetteer data were being used by the end-user analysts.

The need for more inclusive gazetteer data could be summed up by these descriptions sent to us by members of the IC:

> *What would make the biggest difference to us would be to have an integrated gazetteer maintained by the government that consumes all known sources of place-name information served out as a service so that end users do not have to worry about maintenance and upkeep.*
>
> *To support our future and current research, I am interested in having a public domain, unified gazetteer that contains both domestic and foreign names. These names would include names approved by the U.S. Board on Geographic Names (BGN), WordNet names and other names which are not approved but which are commonly used.*

> *We would like to have access to a service that will allow us to pass to it plain language text, IC-MSP metadata tagged text or Geo Location based returns (Fields from a result set of a query that contain Geographic information) and receive in return lat-long values. These lat-long values will then be passed to a mapping service like DGINet and used to visualize the data on a map. We need access to a good Gazetteer service to complete the package.*

We gathered a detailed list of requirements from input by members of the IC through telephone interviews, email exchanges, and in-person meetings. The detailed list of compiled requirements is presented in tabular format in Appendix A.

This detailed list was presented at the May 22-23, 2007, ***Federal Gazetteer Users Group Meeting*** co-hosted by the United States Geologic Survey (USGS), the National Geospatial Intelligence Agency (NGA), and the PATTON Alliance. This meeting was the result of the outreach by the PATTON Alliance on behalf of the IC. There was a need to bring together members of the IC with members of the BGN, the USGS, and NGA. The PATTON Alliance served as the meeting organizer and both the USGS and NGA provided staff and resources to hold this day-and-a-half-long meeting. This was the first time that members of the IC had a chance to present the need for an integrated gazetteer with both foreign and domestic geographic place names to both the USGS (is the agency responsible for the domestic geographic names) and the NGA (is the agency responsible for the foreign place names). The complete meeting minutes are in Appendix B.

This venue served as the ideal place to validate the requirements that we had been gathering. The list of requirements we created was presented to the meeting attendees and we received feedback at the meeting. No major requirements from the IC were missing. From this list we derived the following common requirements from members of the IC:

1. The need for the data to be available free or at low cost to government agencies.

2. The need for a set of gazetteer data that included both domestic and foreign names that were structured the same way and truly integrated.

3. The ability to ingest and incorporate gazetteer data from a wide variety of sources in a wide variety of languages and scripts.

4. The need to include local/colloquial place names that are not officially recognized by the US federal government.;

5. The need for variant spellings, variant references to the same place, abbreviations used, and adjective forms of a place name (i.e., Chinese Coast, Iranian Border, etc.).

6. The ability to manage historical names as well as changes in names over time with time periods the name was used.

7. The need for ancillary data that are not normally part of a gazetteer, but would help in disambiguation of a place name, such as population data, variant spellings, regional

names, ethnicity, and religious affiliation, and data from the Central Intelligence Agency (CIA) Fact Book and WordNet.

8. The ability to rapidly update place-name information in the fielded systems by automated methods with very little human intervention.

9. More geographic name content, especially in areas of emerging threats.

10. Access to a gazetteer service, rather than to a gazetteer data set. Some users indicated that maintaining gazetteer data was too costly in time and expertise, thus they would rather connect to a service that would maintain these data.

11. The ability to create different types of output for use in a variety of systems, such as map display with highlighted text, export tab or coma delimited file of the references and coordinates, export KML so it could be used in conjunction with Google Earth, and export common GIS formats such as Shapefiles for use in other GIS systems.

12. Be standards-compliant (ADL gazetteer standards or Open GIS Consortium) if it is without penalty to speed in data retrieval, accuracy, or system performance.

3. COMPILATION OF GAZETTEER SOURCES FROM THE INTERNET

One of the activities that we undertook to better understand the current gazetteer environment was to perform a survey of available digital gazetteers, thesauri, and other gazetteer-like data sources. We did an open source web-based search for additional sources of gazetteers and gazetteer data and grouped the information in logical categories. Based on the data available from the websites, we compiled a table of information, listing sources, URLs, data sources, standards used, and any comments. The categories we used were:

- **Government Gazetteers** - The USGS GNIS and the NGA GNS.
- **US Government-Funded Research on Gazetteers -** MITRE Inc. IGDB.
- **Research Gazetteers** - ADL, Getty Thesaurus of Geographic Names, Electronic Cultural Atlas Initiative, and others.
- **Commercial Gazetteers -** MetaCarta, Columbia Gazetteer of the World.
- **Web-based Gazetteers -** Geonames, Falling Rain Gazetteer, Information Commons Gazetteer, Worldwide Index.
- **Country-specific Gazetteers -** EuroGeonames, Qatar, Electronic Cultural Atlas Initiative – Iraq.

None of these gazetteers met the stated needs or requirements that we were given. Most of these gazetteers incorporated both the USGS GNIS and NGA GNS as their primary sources of information. Many of these gazetteers have additional place-name content; however, just having more names did not meet the requirements. Many of these gazetteers had licensing issues and were not freely available. An additional concern for web-based gazetteer information was the inability to determine the source of the data. The complete table of sources is in Appendix C.

4. MITRE'S INTEGRATED GAZETTEER DATABASE

The PATTON Alliance was specifically directed by NGIC to evaluate the MITRE IGDB as a possible solution for a more integrated gazetteer solution for the text-based search tool, *GeoLocator***.** The IGDB was designed to support the question-answering systems of the Advanced Research and Development Activity (ARDA) Advanced Question Answering for Intelligence (AQUAINT) program (Mardis, 2005). The aims of this prototype integrated gazetteer were to (Mardis, 2005):

- Combine official comprehensive gazetteers of worldwide places.
- Aggregate names in both their original foreign-language form as well as Latinate transliterations.
- Provide a schema that provides a common view of the disparate sources from which it is drawn.

The IGDB was originally created from four publicly available gazetteers: the NGA GNS (NGA, 2007) for foreign place names, the USGSGNIS (USGS, 2007) for domestic names, and two other U.S. government sources: the CIA World Fact Book (CIA, 2007) and TIPSTER, a U.S. government-sponsored text research program (Irie and Sundhiem, 2004). In addition, two small foreign language test data sets were derived from other sources, Arabic-English and Chinese-English (Mardis, 2005). These two small test data sets were used to demonstrate how the IGDB could be used to store place-name information from other sources as secondary sources.

The IGDB was designed as an integrated gazetteer that could consume gazetteer information from both primary sources (like the GNS, GNIS, and the CIA Fact Book) and secondary sources (like the foreign place names that are linked to data from the primary sources). The IGDB was designed not just as a cross reference to information from many sources, but with a schema that mapped all entries to a uniform schema based on the ADL (ADL, 1999) gazetteer feature-type schema for place names (Irie and Sundheim, 2004). This in essence created a separate database table that contained unique ID numbers and feature numbers as well as information that tracked the origin of the data source. After the data were loaded into this integrated table, duplicate and erroneous entries were removed and the tables were normalized (Irie and Sundheim, 2004; Mardis, 2005). As part of this design, an updating mechanism was designed that allowed for rapid update of the integrated gazetteer as updated source material became available (Irie and Sundheim, 2004; Mardis, 2005). The IGDB was also designed to access auxiliary information such as population, language, etc. in a set of related tables that tie to the primary information as secondary sources (Mardis, 2005).

It is this integration of multiple sources in multiple languages and the rapid update capability that makes the IGDB a good candidate gazetteer to incorporate into fielded information systems.

5. GAP ANALYSIS

The last step in our process was to perform a gap analysis between the requirements gathered by talking with members of the IC, by a qualitative review of the IGDB (Irie and Sundheim, 2004; Mardis, 2005), and through discussions with Scott Mardis, the technical lead for the IGDB at MITRE.

As a result of the May 22-23, 2007, ***Federal Gazetteer Users Group Meeting***, several working groups were formed. These working groups will be chaired by NGA and will address many of the issues identified in the gap analysis below. The working groups formed at the ***Federal Gazetteer Users Group Meeting*** will be covering these five topic areas:

1. ***Content –*** This working group will address issues concerning content of a gazetteer that would be more useful for the intelligence community. Examples are:

    - Search ability.
    - Unique name ID.
    - Name with and without diacritics.
    - Single-feature classification in GNS and GNIS.
    - Point and bounding box for coordinates.
    - Position in hierarchy.
    - Multilingual.
    - Multilingual script.
    - Phonetics.
    - Official abbreviations.
    - Adjective forms.
    - Temporality (time frame for names)
    - Various file formats
    - Map files
    - Image Files
    - Data sources for measure of reliability.
    - Expanded feature classes (e.g., critical infrastructures).
    - Population data.
    - Elevation.
    - Reliability/accuracy measures.

2. ***Standards –*** This working group will consider issues related to standards for gazetteers and ***"place-name intelligence***". a phrase coined by Randall Flynn .the chairman of the BGN. Some of the standards issues that were brought up by the meeting participants include:

    - Unicode (use it).
    - Transliteration (use these standards more).
    - Standard gazetteer data model needed.
    - Data-exchange formatting standard.
    - Glossary of terms.

3. ***Integration –*** This working group will address issues related to the integration of both the foreign and domestic geographic names, whether it be through software, hardware, or user interfaces:

    - Merge GNIS and GNS.
    - Foreign official datasets.
    - Normalization.
    - Importing and exporting data.
    - Conflation tools.
    - Abstractions to allow integration.
    - Software services to assist in integration.
    - How to integrate from “shoe boxes” (distributed model).

4. ***Interface and Access –*** This working group will consider issues relating to how end users access and use the data:

    - Web query.
    - Multiple-format downloading.
    - Web services.
    - Mirroring.
    - Public and classified access.
    - Findability (how easy/hard to find the gazetteer data/service).
    - Usability (define classes of users).
    - Interface to the Help file.

5. ***Other –*** This working group will work some of the larger over-arching issues, such as:

    - The need for use cases from members of the IC.
    - Data collection from non-traditional sourcesand the reliability of these sources.
    - Prepare for a BILLION place-name entries in the near future.
    - Archiving.
    - Metadata.
    - Data conversion from other sources.
    - Web harvesting.

Much of the work that will be done by the working groups will contribute to closing the gaps identified in the gap analysis performed for NGIC for the suitability of the Integrated Gazetteer Database for NGIC. The gap analysis is summarized in the table below:

Table 1. Gap Analysis

Note: Items in ***dark italic*** are related to the working groups established as part of the ***Federal Gazetteer Users Group Meeting.***

Requirement	IGDB Capabilities	Identified Gap	Discussion/Comments
1. The need to be available free or at low cost to government agencies			IGDB was paid for by US government research funds.
2. The need for a set of gazetteer data that included both domestic and foreign names that was structured the same way and were truly integrated.;	Core function of IGDB.		IGDB contains GNIS, GNS, TIPSTER, WORDNet, CIA Fact Book and can ingest data from other sources. ***There will be an NGA working group established to address this issue.***
3. The ability to incorporate gazetteer data from a wide variety of sources in a wide variety of languages and scripts.	Core function of IGDB. Transliteration of foreign scripts has been demonstrated with the IGDB; it is still under development.	Each new source will need to be analyzed and appropriate tables created.	IGDB contains GNIS, GNS, TIPSTER, WORDNet, CIA Fact Book and can ingest data from other sources. Modifications are on a case-by-case basis. Work will need to be done for foreign scripts and transliteration. Transliteration is a difficult issue for any non-Latinate database. ***NGA has current research going on in this area.***
4. The need for local/colloquial place names that are not officially recognized by the US federal government.	IGDB contains some information – depends on sources.	More local/colloquial names can be added.	New sources need to be investigated. ***The NGA is working to develop a way to obtain field-collected data from the War Fighter in a new program called "Intellimapia-transactional gazetteer."***
5. The need for variant spellings, variant references to the same place, abbreviations used, and adjective forms of a place name (i.e., Chinese Coast, Iranian Border, etc).	The IGDB demonstrated that this is possible.	Although this is available in the IGDB, it still is under development.	Although this is available in the IGDB, it still is underdevelopment. ***The NGA is working to develop a way to obtain field collected data from the War Fighter in a new program called "Intellimapia-transactional gazetteer."***
6. The ability to manage historical names as well	Core function of the IGDB.	The IGDB structure can	***Both the NGA GNS and USGS GNIS have the capability to***

Requirement	IGDB Capabilities	Identified Gap	Discussion/Comments
as changes in names over time.		manage historical names and changes over time.	***store historic names. However the lack of ability to access the information in the current data structure make it of limited use.***
7. The need for ancillary data not normally part of a gazetteer but that would help in disambiguation of a place name: population data, variant spellings, regional names, ethnicity, religious affiliation, and data from the CIA Fact Book and WordNet.	The IGDB handles this by creating secondary tables that are linked to the primary source tables.	Some of these data are not readily available.	Additional sources for this type of information will need to be identified.
8. The ability to rapidly update place-name information in the fielded systems by automated methods with very little human intervention.	Core function of the IGDB.	Update process is not a standard deliverable of the IGDB. The updated functions are not well documented.	The update functionality will need to be completed prior to deployment to fielded systems. ***The NGA is working to develop a way to obtain field-collected data from the War Fighter in a new program called "Intellimapia-transactional gazetteer."***
9. More geographic name content, especially in areas of emerging threats.		Not part of the IGDB.	This is a larger issue – one of data-gathering, not specific to any particular gazetteer.
10. Access to a gazetteer service, rather than to a gazetteer data set.		Not part of the IGDB.	This is a larger issue of specific delivery mechanisms, as well as "ownership and maintenance of gazetteer data. ***Both USGS and NGA are working on such services.***
11. Be standards-compliant (ADL gazetteer standards or Open GIS Consortium standards), if it is without penalty to speed in data retrieval, accuracy, or system performance.	IGDB follows the ADL standard for feature type and UNICODE standards for standardized names.	MITRE did not investigate the use of the Open GIS Consortium Standards. MITRE felt these standards were too complex for their gazetteer task.	Open GIS Consortium Standards are still evolving and may or may not be applicable to this task. Will need to be investigated. ***There will be an NGA working group established to address this issue.***

To deploy the IGDB to a fielded system such as the *GeoLocator* the following work will need to be completed:

- The update capability, which is a core function of the IGDB, will need to be documented and coding will need to be completed so that it can run as a stand-alone application.

- The core databases, the NGA GNS and the USGS GNIS will need to be refreshed with data updated through 2007.

- The IGDB structure can handle local/colloquial names; however, no public sources for this information have been identified. The addition of local/colloquial names will need to be coordinated with the NGA and USGS.

- MITRE demonstrated how the IGDB is a flexible framework that can incorporate place names in Arabic and Chinese. MITRE indicated that foreign languages were a part of the IGDB that was not fully developed and would need additional work before it could be released.

The IGDB meets the majority of the requirements revealed to us through the requirements-gathering process. We feel that the IGDB is a good candidate for deployment within fielded systems such as the *GeoLocator* and that it could be a partial solution to gazetteer needs within the IC.

Integration Costs

The IGDB was released to NGIC in July 2007. NGIC has the authority to release the IGDB to Digital Reasoning Systems, Inc. (DRS) for the *GeoLocator*. The IGDB was delivered to NGIC in its original state, without any modifications, due to funding issues. We discussed with DRS how they could use the IGDB in its current state without any modifications (Mancini, 2007):

- At a minimum, the IGDB contains additional place-name content the *GeoLocator* does not contain, mainly the domestic place names, and the small subsets of Arabic and Chinese place names.

- The IGDB documentation has provided the team at DRS with insights on how to better structure the gazetteer data within the *GeoLocator* to take advantage of hierarchical relationships (a city is within a state which is in a country…).

Some of the requirements not addressed by the IGDB are related to issues concerning the nature of gazetteer data:

- There is not a single standard for gazetteer data (although the ADL and the OGC have proposed some standards that apply to gazetteers and gazetteer data(ADL, 1999; OGC, 2007),

- Collecting and maintaining gazetteer data are expensive and time-consuming tasks. No one user in the IC has the ability or expertise to collect and maintain gazetteer data. The

IC is looking to other federal government agencies, such as the NGA and USGS, to develop and maintain gazetteer data for their use.

- Many in the IC would prefer to connect to a gazetteer service or a web feature service owned and maintained by another federal agency rather than maintain a set of gazetteer data within their own systems.

6. ADDITIONAL ACCOMPLISHMENTS

As part of our research work, additional accomplishments were made by the PATTON Alliance on behalf of gazetteers for the wider IC. Specifically, the PATTON Alliance played the role of an "honest broker" to bring parties together to talk about common issues and map future steps forward. These accomplishments are listed below.

1. The PATTON Alliance Gazetteer Project held a series of meetings in February 2007 that introduced members of the IC who had a common interest in improved gazetteers to one another. Out of this grew conversations between the foreign and domestic geographic names committees (hosted by NGA and USGS, respectively) and the realization that these gazetteer data were being used very extensively by the IC for our current war efforts.

2. The PATTON Alliance Gazetteer Project was able to connect members of the IC who had specific names issues and specific issues with gazetteers to the appropriate parties within NGA.

3. The PATTON Alliance Gazetteer Project co-hosted, the ***Federal Gazetteers Users Meeting,*** May 22-23, 2007, with both the USGS and NGA. This meeting brought together members of the IC with both NGA and USGS and the BGN – both domestic and foreign names committees. Out of this meeting was the formation of five working groups focusing on gazetteers and ***"place-name intelligence***," as Randall Flynn, Chairman of the BGN refers to this work. These working groups, chaired by a person from NGA, will cover these topics: Content, Standards, Integration, Interface and Access, and Other.

4. The PATTON Alliance Gazetteer Project is hosting an SharePoint site for the members of the Federal Gazetteers Working Groups to get started, referencing the materials developed at the Federal Gazetteers Users Meeting, and posting information compiled as part of the Gazetteer Project.

5. The Patton Alliance Gazetteer Project played a role in getting members of the IC who have collected new place-name information out in the field in direct contact with the ***Intellimapia/geowiki*** transactional gazetteer that the NGA is standing up to collect local/colloquial names that is only available to the IC on classified networks.

7. SUMMARY AND RECOMMENDATIONS

The PATTON Alliance Gazetteer Project used an innovative approach by matching existing government-funded research with real IC needs. We evaluated and analyzed the needs and compared it to the IGDB. We obtained feedback from the IC concerning applicability to other IC users. We are in the process of identifying additional gaps and finalizing a plan to modify the IGDB for eventual adoption by the IC as a partial solution.

Some of the unique contributions to both gazetteer research and to the IC in need of gazetteer data that are a result of the PATTON Alliance Gazetteer Evaluation Project include:

- The PATTON Alliance was able to identify key government research efforts regarding gazetteer data.

- We verified one gazetteer research effort, the MITRE IGDB that could be modified and deployed to the *GeoLocator* system.

- We identified key federal agencies that need to be involved in future gazetteer work (USGS, NGA, etc.).

- We identified existing gaps between current IC gazetteer needs and existing gazetteers.

- We brought critical stakeholders together with the USGS and NGA at ***The Federal Gazetteer Users Meeting***, May 22-23, 2007, during which many IC stakeholders were able to discuss their needs for better gazetteer data and to give input to the USGS and NGA on future development efforts.

- We identified ways to close the gaps between needs and existing gazetteers, as well as a path forward to implement improvements in gazetteers for the IC by establishing gazetteer working groups chaired by NGA. These working groups were the outcome of the ***Federal Gazetteer Users Meeting***.

Recommendations

- For the gazetteer working groups to be successful, they will need some logistical support, similar to what the PATTON Alliance provided for the ***Federal Gazetteer Users Group*** meeting in May 2007. PATTON's role as an "honest broker" and meeting facilitator brought together members of the IC and the BGN for the first time. This meeting highlighted some urgent needs that were not previously understood.

- Often at first-time meetings, the energy is high, but when participants return to their normal jobs, the ability to stand-up new work items in the form of a new working group becomes hampered. The important progress made by the PATTON Alliance in improving gazetteers for the IC is at risk if the working groups do not materialize and continue for at least three years. Three years is the minimum amount of time it takes working groups to form, work on an issue, and produce results. The PATTON Alliance could play a role in helping to support these working groups.

As an example, in July 2007, the PATTON Alliance held a short telephone conference call meeting with members of the IC and NGA to provide input into the "Intellimapia Transactional Gazetteer" Project under development at NGA (Hagan, 2007). The outcome of this meeting was that field-collected place-name data from the war fighter has now been delivered to NGA as part of a pilot project for the "Intellimapia Transactional gazetteer" and having the direct involvement of the war fighter community is going to be a critical component of the success of this new form of gazetteer. If funded, the PATTON Alliance could provide this level of support for all of the working groups.

The cost for this effort would be approximately 1 FTE and support for significan travel and meeting expenses, for three years. This would provide the logistical support to establish the working groups and host an annual Federal Gazetteer Users Group Meeting, in which direct feedback from the IC would give the USGS and NGA input on improving gazetteer data and services.

8. REFERENCES

Alexandria Digital Library (ADL) Gazetteer, 1999. Santa Barbara, CA: Map and Imagery Lab, Davidson Library, University of California, Santa Barbara. Copyright UC Regents. http://www.alexandria.ucsb.edu/gazetteer (ADL Gazetteer Development page with links to various clients and protocols that access the ADL Gazetteeraccessed 10/30/06).

Central Intelligence Agency (CIA), 2007. The World Factbook. https://www.cia.gov/library/publications/the-world-factbook/index.html (accessed May 2007).

Hagan, Brian (2007). National Geospatial Intelligence Agency. Personal communication with the author.

Irie, R. and Sundheim, B., 2004. Resources for Place Name Analysis. In *Proceedings of the 2004 International Conference on Language Resources and Evaluation - LREC2004*, pp. 317–320.

Linton, Frank, 2005. "Accelerating GEOINT Analysis: User Feedback to Improve Geospatial Place-Name Services," MITRE Corporation White Paper.

http://www.mitre.org/work/tech_papers/tech_papers_05/05_0247/index.html (accessed May 10, 2007).

Mancini, Peter (2007). Personal communication with the author.

Mardis, Scott (2007). Personal communication with the author.

Mardis, Scott and Burger, John, 2005. Design for an Integrated Gazetteer Database: Technical Description and User Guide for a Gazetteer to Support Natural Language Processing Applications. MITRE Corporation White Paper .http://www.mitre.org/work/tech_papers/tech_papers_06/06_0375/06_0375.pdf

National Geospatial Intelligence Agency (NGA), 2007. GEOnet Names Server (GNS). http://earth-info.nga.mil/gns/html/index.html (accessed May 10, 2007).

Open Geospatial Consortium (OGC), 2007. Specifications and standards. http://www.opengeospatial.org/standards (accessed, May 10, 2007).

United States Geological Survey (USGS), 2007. Geographic Names Information System (GNIS). http://geonames.usgs.gov/index.html (accessed, May 10, 2007).

APPENDIX A: COMPILATION OF GAZETTEER USER REQUIREMENTS

Topic	Requirement/Issue
Cost	1. Gazetteer data must be available to federal agencies at no cost. 2. We use both the GNIS and GNS data in our system. We are considering using MetaCarta, but it doesn't quite meet our needs and it costs too much.
Sources	Open source is better.
	We use both the GNIS and GNS. We are considering using MetaCarta, but it doesn't quite meet our needs and it costs too much.
	We use both GNIS and GNS. The databases are uniquely structuredand not very compatible or amenable to integration.
Normalized Data Base	Want as normalized a database as possible.
National Geospatial Intelligence Agency (NGA)	What is NGA doing to address Gazetteer issues for the IC?
	Which federal agency (NGA?) will maintain the (integrated) gazetteer? We do not have the staff, expertise, time, or money to maintain a gazetteer.
Data Pre-processing	
	NGA data requires “a lot of clean-up” prior to use in text-based search systems
	LOTS of pre-processing of NGA/USGS gazetteer data to make text searching against the data possible
Data Issues	
	Combining the GNS and GNIS into one single gazetteer would be ideal. That way unique IDs for individual features could be maintained over time.
	Having a unique ID for features and feature names that did not change is important. One ID per name/feature. (Currently GNIS and GNS have different feature ID coding.)
	In GNS Historical Names have special characters that prevent text searching, i.e., double parenthesis around historical names.
	Hierarchy is critical for text-based searching. NGA uses Admin. GNIS only points to one. The Getty Thesaurus has a hierarchy for natural feature, which is a good example.
	NGA data iare not complete in the attributes; not enough metadata to assist in text queries. Would like the content of the CIA FactBook.
	No odd characters within fields. NGA data has multiple brackets, quotes, etc.
	No redundant entries.
	NGA feature codes - NGA data had many feature codes that do not meet project objectives.
	NGA data deficiencies: Null values found within the following fields: DIM - dimension- elevation or population data; PC - Populated place classification; ADM2 – Second-order administrative division; LC - language code; Short form - a specific part of the name that could be substituted for the full name; Generic- The descriptive part of the full name (does not apply to populated place names).

	"Augment the database with entities that have the highlighted fields below, as well as entities that populate the schema elements previously identified: lower threshold for population levels in populated places, if there is no population designator. There is a need to be able identify small villages. Elements highlighted from the NGA GNS: GNS_ID; LAT; LONG; DSG; CC1; NT; Full_name_nd; Country_Capital.
	Data structure is problematic – EX: In Afghanistan NGA uses Provence as part of the name, which doesn't help in searching.
	LOTS of pre-processing of NGA/USGS gazetteer data to make text searching against the data possible.
	Continent names are missing.
	GNIS compound structure does not match how text-based search engines use the data. E: Washington, state of
	Sometimes there is attribution included as part of the name that must be stripped out.
	There are non-letter characters in many fields that must be stripped out.
	Administrative-level generic is included as part of the name. EX: Ontario Provence instead of just Ontario. Mountains are always problematic -- Mount X, X Mountain, X Peak, etc.
	Lakes are problematic - Lake X, X Lake, etc.
	Valleys are hard to find. Since the real thing is around a river, not something named as a valley, analysts end up searching for an area around a river.
Metadata	Complete metadata - NGA data did not have enough data within the file that would assist in the textural queries.
	We need complete metadata concerning the gazetteer entries (identifying the original source, changes/update history, etc.).
DATA UPDATE	
	The gazetteer data should be easily updated.
	Updates must be easy and as automated as possible. When new releases of the data from the authoritative sources, it must be easy to consume/use the new data. No manual examination of the data should happen.
	Refreshing the data is problematic because of the pre-processing that must be done for the data to be used within a text-based search tool.
	We do not want to be the owner or maintainer of a Gazetteer; rather we would prefer accessing gazetteer data through a web mapping service.
	We need the ability to track change/updates to place names entered into the gazetteer.
HISTORICAL NAMES	
	Historical place names are important. A time period associated with a name is important.
	Historical Names: Historical names should be kept within the system. "I want to find Bombay, India, for instance, even though the name has changed."
	In GNS Historical Names have special characters that prevent text searching (double parenthesis around historical names).
	We need a way to manage historical place names and changes to that place name over time.

ENHANCEMENTS	
	Identify a design that allows disambiguation of location and place names. Provide additional Information that describes location from a variety of points of views.
	Identify a design that allows disambiguation of location and place names: official designations, colloquial designations.
	Combining the GNS and GNIS into one single gazetteer would be ideal. That way unique IDs for individual features could be maintained over time.
	Single hierarchy structure between GNIS/GNS – both are very different and it would be the best if GNS and GNIS had the same database structure.
	Need a Unique ID. The Unique ID then allows links to other spatial representations - i.e., Name of the State of Virginia, could then link to a spatial representation (coordinates) and or map of the state of Virginia. Linking of spatial products based on unique ID.
	Need geographic regions – Sahara Desert, Great Plains, etc.
	Need both formal and informal administrative areas.
Population Data	Additional data: Population down to small villages.
	Need population data in broad categories. On a world level – need to have population data for places larger than 15,000 people. In the US, need population data for places greater than 5,000 people.
Variant Names/Spelling	Variant names, need a way to get contextual information concerning the name.
	"Add missing place names (smaller cities, neighborhoods) and variants (i.e.., as found in the popular press). “We get asked about many place names that are not in Geonames. We need all the place names and their variants. We depend on Geonames as the authoritative source, but Geonames doesn't always have what we need."
	Abbreviations – Would like known abbreviations available in the gazetteer.
	Need adjective form place names, i.e., China, Chinese; Iran/Iranian. For example, Chinese coast – very helpful clues for context.
Name use caveats	Would like information on caveats for name usage.
Native language	"A view of foreign names in their native language."
	Recognized soundex /phonetic spellings for the spellings of names. This would be very helpful in text-based searching.
	Need place names in both native language and with the name variants for the same place name.
	We need information on the ethnicity of a place name.
Regional/Colloquia Names	"Additional Gazetteer information needed could include such items as regional (colloquial or tribal) additional reference names for the same place. Such data could be utilized for multiple name resolution or conflation purposes."
	Local/tribal names are important. Also local/tribal regions are important.
Additional Content	Additional Names needed: Middle East, Africa, SE Asia.
	Additional Gazetteers: Interested in consuming additional gazetteers from other sources and other countries.
	Names of the people: Chinese for China, Taiwanese for Taiwan, Tico for

	Costa Rica, etc.
	We need ways to access additional non-location-based information, such as the CIA Fact Book, TIPSTER, WORDNet.
Disambiguation/Contextual information	
	Context information: Need a way to 1) distinguish between multiple points for the same location and 2) distinguish between multiple locations for the same name.
	Automate disambiguation that the analyst currently performs manually; use fielded data to help disambiguation of other fields in addition to information identified through unstructured text analysis.
	Need information for context for location disambiguation. The village X near Kabul. The analysts would search for all villages near Kabul.
	Containment: The gazetteer database should have a way to link a place name to political and administrative areas, e.g., place X is located in Y county/township, which is in A province, which is in B country, which is located on C continent.
	Co-refererence: If multiple sources contain information about a common place, the gazetteer database shall maintain multiple place entities in the database.
	If there are multiple returns of a place name in one or more locations, there must be a way to disambiguate or narrow down the possible choices. This should be as automated as possible.
QA/QC	
	"…Requirements on the underlying gazetteer are stringent on quality of data and performance. Additional gazetteer entries need to maintain the same scrutiny of data quality prior to incorporation."
	We already send NGA information on data that is incorrect.
	How do you send in corrections to NGA?
STANDARDS	
	Be standards-compliant (Alexandria Digital Library gazetteer, OGC, ISO, etc.) if it is without penalty – of speed in data retrieval, accuracy, or system performance.
	OGC Standards: Not really using them now in geo text work. Standards are good and we want to move toward it.
OTHER ISSUES	
Non "Official Place Names"	"Add missing place names (smaller cities, neighborhoods) and variants (e.g., as found in the popular press). "We get asked about many place names that are not in Geonames. We need all the place names and their variants. We depend on Geonames as the authoritative source, but Geonames doesn't always have what we need."
	Local names not recognized by the US Government: For example, Jerusalem. The State Department does no recognize Jerusalem as being part of any Nation State. It adheres to a concept of Internationalization for the city and surrounding areas. Prior to this Administration, we did not even recognize nearby Israeli Settlements.
	The USGS GNIS and the NGA GNS are recognized as "authoritative sources." However, this is from the US government's point of view and limits place names to those accepted by the US government. This gazetteer should be able to contain and manage place names that are recognized by other global entities such as the United Nations (UN) and other countries.

Searching strategies	Many analysts take a name as an end user might refer to it and then add variants that might be used to search for places.
	Need information for context for location disambiguation. The village X near Kabul. The analysts would search for all villages near Kabul.
	Would like the ability to search in foreign languages.
	Would like the ability to search using phonetics.
	Would like a more robust wild card search capability.
	Would like the ability to use “Fuzzy' Search” logic.
Output	Different types of output needed: • Map display, highlighted text. • Export a tab delimited file of the references and coordinates. • Export KML so it will pop up using Google Earth, using the free version of Google. • XML output – All of the original text with names tagged and gazetteer information for the names, which become geospatial data . • Would love to create an intel standard for the XML to make the information more sharable between agencies and vendors.They touch the open internet so that they can display some information on Google Earth.
	Would like to take retrieved place-name data and display it on Google Earth.
Use of Web Feature Services	
	We would like to have access to a service that will allow us to pass it plain-language text, IC-MSP metadata-tagged text ,or GeoLocation-based returns Fields from a result set of a query that contains geographic information) and receive in return lat-long values. These lat-long values will then be passed to a mapping service like DGInet and used to visualize the data on a map.
	We do not want to be the owner or maintainer of a Gazetteer; rather we would prefer accessing gazetteer data through a web mapping service or web feature service.
	We need access to a good Gazetteer service to complete the package.

APPENDIX B: MEETING NOTES OF THE FEDERAL GAZETTEER USERS MEETING MAY 22-23, 2007

Meeting Notes of the Federal Gazetteer Users Meeting
May 22-23, 2007
US Geological Survey Headquarters, Reston, Virginia

Notes edited by Denise Bleakly Sandia National Labs.

The **Federal Gazetteer Users Meeting** was held at the US Geological Survey Headquarters in Reston, Virginia, on May 22-23, 2007. It was hosted by the US Geological Survey (USGS), the National Geospatial Intelligence Agency (NGA), and the PATTON Alliance (Sandia National Laboratories).

Representatives from various users groups included members of the Intelligence Community (IC) and the Library of Congress (LoC) who met with representative of the NGA and USGS to discuss the future needs for gazetteers, place-name services and "place-name intelligence".

Day 1, Tuesday, 22 May 2007

Summary:
Day 1, involved introductory speeches and welcomes from the host organizations of the meeting: The PATTON Alliance, the USGS, the NGA, and the US Board on Geographic Names, then we heard from members of the IC about their needs for improved gazetteers and gazetteer services.

Welcomes
Denise Bleakly, Gazetteer Project Lead of Sandia National Laboratories, opened the Meeting. She is with Sandia National Laboratories which is home to the PATTON Alliance, whose acronym means "Preparing Analytic Tools and Technology for an Operational Network." PATTON is a technology middleman to get *"government-funded technology ready for integration and deployment to specific end users,"* which in this case means the US government.

Denise Bleakly stated in the written agenda that the purpose of the conference was *"to bring together PATTON Alliance Members, with NGA, USGS and other federal stakeholders to discuss issues concerning the use of gazetteers in support of the Intelligence Community (IC), to identify needs and explore how to close the gaps between user needs and current gazetteers."*

The goals of the conference were the following:

1. Understanding the role that NGA and USGS play in the identification of geographic place names.
2. Identifying the gaps between a government-funded research gazetteer and what is needed in a fielded government system.
3. Determining if a PATTON-identified technology will meet the needs of fielded federal government systems.

The NGA welcomed the conference participants and listed the goals of the conference:

1. The opportunity to exchange information.
2. The discussion of future trends for gazetteers.
3. A forum for the needs of gazetteer users.

NGA hoped that information gaps in gazetteers based on users' evolving needs will be filled and that such conferences will be held annually.

Lou Yost, USGS and Executive Secretary of the Domestic Names Committee (DNC) of the BGN, welcomed everyone to the US Geological Survey (USGS). He mentioned that the conference will be tilted toward foreign place-name gazetteers. His agency, the USGS, produces the Geographic Names Information System (GNIS), which is the official digital gazetteer of US place names.

The PATTON Alliance Program Manager at Sandia National Laboratories, welcomed conference attendees on behalf of the PATTON Alliance. She stated that members of the PATTON Alliance are technology developers reviewing the transition of new technology from research to deployment. The Gazetteer Project is aimed at searching for existing technology and determining current needs for gazetteers. Current funding for their gazetteer evaluation project ends in July 2007.

Randy Flynn, NGA and Executive Secretary of the Foreign Names Committee (FNC) of the BGN, represented the BGN.

Randall Flynn gave an interesting context-setting presentation. In it he pointed out that they are really in the business of ***place-name intelligence***. Their users' expectations are high:

- Adherence to standards.
- Ease of access.
- Toponymic information that meshes seamlessly with other information.
- The right content.

Their past successes have occurred in mathematical space. Their new challenge is precision identification in human space.

The challenges facing them have to do with:

- Scale.
- Diversity of content.
- Colloquial toponymy.
- Ambiguity.
- Deeper analytic capabilities.

Their services need to include:

- Non-English, non-Roman alphabets.
- Non-textual data.
- Place-name trolling.
- Conflation/deconfliction. / deconflation
- Contextual confidence indicators.
- Virtual collaborative collection.

In addition, he discussed the traditional and historical use of gazetteers and that users have expected the BGN to gather and compile names. He brought up the concept of ***place-name intelligence,*** which brings new challenges for future gazetteers.

The Deputy Executive Secretary of the Foreign Names Committee (FNC) of NGA, gave an overview of the role of the Board on Geographic Names (BGN). He encouraged users to offer their research and advice to BGN to improve and expand gazetteer data.

Here are some of the challenges that the FNC faces:

- Resistance to change.
- Lack of sources.
- Staff shortage.

- Larger scope of gazetteer needs.

The Executive Secretary of the of the Domestic Names Committee (DNC) of USGS, discussed the role of the DNC. Members of the DNC include representatives from the following agencies:

- Department of Agriculture.
- Department of Commerce.
- Department of the Interior.
- Department of Homeland Security.
- Government Printing Office.
- Postal Service.
- Library of Congress.

The committee goes over approximately 400 name proposals per year and enters approximately 3000 names into the names database (GNIS) per month.

There are 50 state names authorities that form the Council of Geographic Names Authorities (COGNA).

The, USGS, introduced the topic of *The National Map*, which covers the entire country with seamless USGS maps. The names come from the GNIS, which contains over 2 million features.

NGA, introduced the future GNS that stores the foreign-names database of the US government. They mentioned that the future GNS is an Oracle database rather than a Sybase database. The future GNS is part of the GGMA initiative for an integrated data-centric environment (i.e. a conflation of digital databases).

NGA, then demonstrated the future GNS website.

NGA, discussed future trends in gazetteer production and maintenance. They are working to bring automated processes to ease the work that has been mostly entered manually and, therefore, is labor-intensive and time-consuming. Among the automated processes are the following:

- Optical Character Recognition (OCR) for lists of names.
- Transliteration.
- GeoDDupe, which is a gazetteer conflation tool being worked on by the University of Maryland.
- MITRE's Geospatial Data Integration Algorithms Tool for gazetteer file matching.
- NGA's Intellimapia – NGA is working to have a user-contribution website by 2008, in which users contribute names data that is similar to the existing Wikimapia, which consists of over 3.5 million user-contributed names.
-

Issues and Concerns about Gazetteers and Gazetteer Data from the Intelligence Community

Five different representatives from the IC spoke about their needs for gazetteer data and their ability to effectively use the NGA GeoNames Server for foreign place names and the USGS Geographic Names Information System.

Comment 1. The IC analyzes foreign information of interest and needs the gazetteer information to plot place names from its information onto maps. The IC would like to see improved coordinates in gazetteers. A lot of IC gazetteer information comes from the GNS.

Comment 2. The IC is a heavy user of GNS. Unfortunately, many features, such as buildings and streets, are not in the GNS, and, consequently, individual analysts are tasked to do the research of these missing features. An analyst provided input that an improved gazetteer is one that includes named features found on large-scale maps and has precise coordinates. The IC would also like to see the capability of proximity and buffer searches and other ArcGIS functions (i.e. an ArcGIS-based gazetteer).

The IC needs:

- Entities below Admin1.
- Multiple place names associated with a given place.
- Neighborhood details in large cities.
- Street addresses in foreign countries.
- Variant names, multiple languages.
- Place names misspelled, bad transliteration, translated by other nations.
- Large, prominent building in foreign cities.
- Transportation facilities to include ports and other modes of transport.
- More precision in coordinates down to seconds in decimal degrees.
- Proximity search and radius search in GNS.
- Have available a gazetteer within ArcGIS.
- They use Google Earth but mostly ArcGIS.

During this discussion, *Google Earth* as a visualization tool was mentioned, and another visualization tool called *World Wind* by NASA was brought up by members of the audience.

Comment 3: The IC would like to see more attributive data in the GNS, such as elevation and population.

Comment 4: This IC member, emphasized that gazetteer information is constantly needed. His office receives many inquiries from Foreign Service personnel abroad that can be answered by searching in the GNS. The people, however, find the GNS too awkward to use and therefore prefer to call his office. He recommends strongly that the GNS must be easier to use, especially for the unsophisticated user. Codes should be replaced by clear, spelled-out information, such as country names and feature designations. GNIS, which is the US names gazetteer, is easier to use because it is intuitive. The GNS should also contain links to other sources of data so additional data, even photographs, can be found. The GNS search capability in other languages and scripts would also be useful. Also, the returns of a name search should be the most likely match should ranked with the first on the list of returns be the most used or most likely match.

Library of Congress (LoC), spoke about LoC's prototype project, the *World Digital Library*. As do other organizations, the project relies heavily on gazetteer data. Some of their issues about gazetteer data are the following:

- Poor data.
- Inconsistency in metadata.
- Getting data to adhere to standards of foreign libraries involved in the project.
- Data file conflation.

After the presentation by LoC, a discussion started concerning the definition of *neighborhood*. USGS does not use *neighborhood* in its gazetteer, because it is a vague term.

Also, a discussion concerning separate uses of gazetteer data emerged that considered people end users, such as those who user end user analysts, versus the "machine users," for whom a gazetteer data set was needed as part of a larger system for many uses. The "people end-users" needed better user interfaces that were more intuitive and user-friendly, while the "machine end users" needed better data structures, data models, standards, and data dictionaries.

Day 2, Wednesday, 23 May 2007

This day's discussions were about the future direction of gazetteers. The participants were unanimous about the need for gazetteers to improve and fill gaps that users need.

The day was designed to have two speakers, Frank Linton, of MITRE Corporation who did a study of end users' experience of GNS, and Denise Bleakly, and her experience in collecting user requirements for gazetteers for intelligence systems.

After these two presentations, a discussion was facilitated by Judy Moore of Sandia National Laboratories, to list action items and next steps.

Frank Linton, MITRE, wrote a report in 2005 about the GNS and suggestions for improving it. The suggestions came from interviews with many GNS users.

Frank mentioned that the GNS is "painful" to use. He showed the GNS website and explained why it was so. He said that it was difficult to understand the content without a lot of reading and studying. Here are other comments:

- Codes were not intuitive.
- Population data is missing.
- Scale selection is not available.
- Links to additional sites for information are not available.

Frank recommended some changes to the GNS:

- Users should be encouraged to contribute names information.
- News of data updates should be available.
- Dates of names data and their sources should be given.
- Allow users to select the data that they need.
- Offer various formats for downloading.
- Provide phonetic aids.
- Improve wildcard searches.
- Prevent searches to time out.
- Provide alternative coordinate systems.
- Search in foreign script.
- Make GNS easier to find online.

As expected, Frank's presentation evoked a passionate discussion. Here are some points:

- As a start, the user interface should be corrected first before going on to the gazetteer's content and form.
- There needs to be a clear definition of gazetteer. What is its purpose? What should it contain? The gazetteer must evolve to fulfill current and future user needs.
- The GNS is problematic for the user perhaps because it contains foreign data with which users are unfamiliar.
- Provide users a template of a data model so that they can produce their own specific gazetteer.

Denise Bleakly, Gazetteer Evaluation Project Lead, Sandia National Laboratories, presented the PATTON Alliance Gazetteer Evaluation Project. The PATTON Alliance members are looking at the system level to make improvements to the current gazetteers. She mentioned that some foreign governments have their gazetteers online and are good sources of data. The problem, however, is how the data can be integrated with our gazetteers. While striving for the ideal gazetteer, problems must be solved in turn rather than trying to solve all problems before releasing a new gazetteer. The expectations

of the gazetteer have evolved quickly in the high-tech reality, surpassing its original intent of providing a list of names, description, and geographic location.

The PATTON Alliance Project Manager, from Sandia National Laboratories, guided the group discussion further by offering five categories in which we can write ideas for improving gazetteers:

- Content.
- Standards.
- Integration.
- Interface and Access.
- Other.

Content ideas:

- Search ability.
- Unique name ID.
- Name with and without diacritics.
- Single-feature classification in GNS and GNIS.
- Point and bounding box for coordinates.
- Position in hierarchy.
- Multilingual.
- Multilingual script.
- Phonetics.
- Official abbreviations.
- Adjective forms.
- Temporality (time frame for names).
- Various file formats.
- Map files.
- Image files.
- Data sources for measure of reliability.
- Expanded feature classes (e.g., critical infrastructures).
- Population data.
- Elevation.
- Reliability/accuracy measures.

Standards ideas:

- Unicode (use it).
- Transliteration (use these standards more).
- Standard gazetteer data model needed.
- Data-exchange formatting standard.
- Glossary of terms.

Integration ideas:

- Merge GNIS and GNS.
- Foreign official datasets.
- Normalization.
- Importing and exporting data.
- Conflation tools.
- Abstractions to allow integration.
- Software services to assist in integration.
- How to integrate from “shoe boxes” (distributed model).

Interface and Access ideas:

- Web query.
- Multiple-format downloading.
- Web services.
- Mirroring.
- Public and classified access.
- Findability (how easy/hard to find the gazetteer data/service).
- Usability (define classes of users).
- Interface to the Help file.

Other ideas:

- Need for use cases.
- (Who will get these requirements?)
- Data collection from non-traditional sources – reliability?
- Prepare for a BILLION place-name entries in the near future.
- Archiving.
- Metadata.
- Data conversion from other sources.
- Web harvesting.

Work-approach priorities:

- Build working groups for each of the five categories.
- Scan archived and microfilmed names cards for their wealth of additional information.
    - Scan NGA index cards to get information out to the larger world – analysts.
    - MITRE has some technology.
    - BGN index cards?
    - BGN case files?
    - Library of Congress place-name authorities?
- The standards group needs to set up standards that need to change.
- The other groups need to set up requirements. The requirements need to be identified to justify the resources needed for the gazetteer project.
- The PATTON Alliance has a SharePoint site. It can be accessed by anyone with a user ID and password. These are set up by the PATTON Alliance.
- Working groups should hold their first meeting by July 2007. NGA will chair each group.
- The next Gazetteer Users conference should be a full two-day conference held next year in May.

1. It was decided to formulate several working groups to develop requirements and make recommendations to the government for meeting the needs identified in this conference. The plan is to convene the first set of working group meetings before the end of July 2007.

2. Sandia will work with NGA to make sure these first meetings are held and the notes from this meeting are provided to these groups.

3. Sandia will also grant access to the PATTON SharePoint site as a starting point for there working groups. NGA has agreed to provide a leader for each working group.

4. It was also decided to make this conference an annual event. NGA and USGS will take on the responsibility to host this conference as an annual event. They will consider having it in this same

timeframe or placing it adjacent to another larger Geo conference like GEOINT. An opportunity for vendor displays or demos will be included.

APPENDIX C: GAZETTEER EVALUATION MATRIX

Links verified July 2007

Candidate Gazetteer / URL Link(s)	Data Sources	Standards	Comments
US Official Gazetteers			
United States Board on Geographic Names Website, BGN http://geonames.usgs.gov/	N/A	N/A	The U.S. Board on Geographic Names (BGN) is a federal body created in 1890 and established in its present form by Public Law in 1947 to maintain uniform geographic name usage throughout the federal government. The Board comprises representatives of federal agencies concerned with geographic information, population, ecology, and management of public lands. Sharing its responsibilities with the Secretary of the Interior, the Board promulgates official geographic feature names with locative attributes as well as principles, policies, and procedures governing the use of domestic names, foreign names, Antarctic names, and undersea feature names.
Domestic Names Geographic Names Information System, GNIS http://geonames.usgs.gov/domestic/index.html	The U.S. Geological Survey developed the GNIS for the U.S. Board on Geographic Names as the official repository of domestic geographic names data; the official vehicle for geographic names use by all departments of the federal government.	Created own internal standards for domestic names and features.	The Geographic Names Information System (GNIS) is the federal standard for geographic nomenclature. The U.S. Geological Survey developed the GNIS for the U.S. Board on Geographic Names as the official repository of domestic geographic names data; the official vehicle for geographic names use by all departments of the federal Ggvernment; and the source for applying geographic names to federal electronic and printed products.
Foreign Names Geographic Names Database (GNDB) from the GeoNet Names Server (GNS) , NGA http://earth-info.nga.mil/gns/html/index.html	The geographic Names Server is the official repository of standard spellings of all foreign places names sanctioned by the United States Board on Geographic Names.	Created own internal standards.	Source for official foreign geographic place names.
Research Gazetteers			
Alexandria Digital Library (ADL) Project Gazetteer Development http://www.alexandria.ucsb.edu/gazetteer/	USGS Geographic Names Information System (GNIS); NGA GeoNames (GNS).	Yes - Created the Gazetteer Content Standard and	Has created and published a Gazetteer Service Protocol, also has a thesaurus of Geographic Names, 5/2007. The ADL is not actively updated. Data is about 10 years old.

Candidate Gazetteer / URL Link(s)	Data Sources	Standards	Comments
		Relational Database Model for Gazetteer information.	
Getty Thesaurus of Geographic Names (TGN) http://www.getty.edu/vow/TGNSearchPage.jsp	5/2007 This thesaurus was compiled from government sources, information indexing projects, and manual entries from hard-copy sources. The focus of this thesaurus is art, architecture, and material culture.	This is a thesaurus, compliant with ISO and NISO standards for thesaurus construction.	Not a GIS and has limited coordinates. It contains hierarchical, equivalence, and associative relationships.Coordinates are approximate and are intended for reference only. It is not comprehensive.
Electronic Cultural Atlas Initiative (ECAI) http://ecai.org/projects/gazetteer/	An extension of the Alexandria Digital Library Project Gazetteer and applies it to art and culture.	Thesaurus.	Adapted the ADL standards for their work. Main Funded Project -- 2002 "A Multilingual Gazetteer System for Integrating Spatial and Cultural Resources." Not clear if this work is continuing.
Academia Sinica http://www.sinica.edu.tw/main_e.shtml http://pgis.sinica.edu.tw/en/research.html	Unknown.	Unknown.	A Chinese academic group working to produce a gazetteer of Chinese place names. This group is working as a partner in the Electronic Cultural Atlas Initiative (ECAI) A power point on gazetteer work: http://www.sinica.edu.tw/~metadata/bibliography/proceeding/presentation/pnc2002ps_01.pdf#search=%22geographic%20place%20names%22
Commercial Gazetteers			
MetaCarta http://www.metacarta.com/	USGS GNIS; NGA GeoNames (GNS), and proprietary sources.	Unknown.	Has several public research papers available: http://www.metacarta.com/docs/Corporate_White_Paper.pdf http://www.metacarta.com/docs/Public_Sector_White_Paper.pdf http://people.mokk.bme.hu/~kornai/NAACL/WS9/ws910.pdf
Columbia Gazetteer of the World on line http://www.columbiagazetteer.org/	Built on the data contained in the Lippincott Pronouncing Gazetteer, the Columbia-Lippincott Gazetteer of the World, USGS GNIS, and other academic sources.	Unknown.	A user login is required to access the website. The names catalogued are names used by the official nation agencies and transliterated into English. English names are the initial entries; alternative local names follow. Population data is included wherever possible. There are terms of use and licensing restrictions. These can be found at: http://www.columbiagazetteer.org/Gazetteer_License_Agreement.html This licensing agreement would not make it

Candidate Gazetteer / URL Link(s)	Data Sources	Standards	Comments
			possible for use within a gazetteer for government use.
Compusult - WES Gazetteer http://www.compusult.net/cslt_prod_dm_gaz.html	NGA Gazetteer, GNIS.	OGC standards-compliant.	"Compusult's WES Gazetteer is an easy-to-install, standards-based gazetteer with lightning -ast name look-up." -- A fee based gazetteer commercial product.
US Gov. Funded Research on Gazetteers			
MITRE, Inc http://www.mitre.org/	USGS GNIS, NGA GeoNames, TIPSTER, NMSU Arabic Names, Harvard University Chinese Names, context information from CIA Fact Book.	Used the geographic type taxonomy from the Alexandria Digital Libraries project as a unifying taxonomy of place type.	MITRE is doing work on text-based searching and linking gazetteer information. Has a technical article about its Integrated Gazetteer Data Base: http://www.mitre.org/work/tech_papers/tech_papers_06/06_0375/06_0375.pdf
Web-Based Gazetteers			
Geonames http://www.geonames.org/	Is an internet project (not clear on who is sponsoring it) to create a "free" gazetteer based on the "wiki" concept. Many sources are listed on their website. http://www.geonames.org/about.html The data is accessible free of charge through a number of web services and a daily database export. Users may manually edit, correct, and add new names using a user-friendly wiki interface.	OGC WFS standard-compliant.	There are forums, blogs, and message boards for this site. http://geonames.wordpress.com/about/ You can download the data and connect to Web Feature Services using this data at: http://www.geonames.org/export/ a listing of web services using geonames: http://www.nearby.org.uk/geonames-webservices.php?sort=Input The geonames.org geographical database is available for download free of charge under a creative commons attribution license. It contains over eight million geographical names and consists of 6.5 million unique features, including 2.2 million populated places and 1.8 million alternate names. All features are categorized into one out of nine feature classes and further subcategorized into one out of 645 feature codes. Geonames.org is already serving up to over 3 million web service requests per day. Geonames is integrating geographical data such as names of places in various languages, elevation, population, and other data from various sources. All lat/long coordinates are in

Candidate Gazetteer / URL Link(s)	Data Sources	Standards	Comments
			WGS84 (World Geodetic System 1984).
Wikimapia http://wikimapia.org Wikimapia blog: http://wikimapiablog.blogspot.com/	Based on a "wiki" concept, individuals can add place-name information.	Unknown.	WikiMapia is a Web 2.0 project to describe the whole planet Earth. It was created by Alexandre Koriakine and Evgeniy Saveliev, inspired by Google maps and Wikipedia. Data can be viewed using Google Earth. As of July 2007 over 3,000,000 entries.
Information Commons Gazetteer http://www.maya.com/infocommons/index.html http://www.maya.com/web/what/papers/maya_infocommons_gaz.pdf	Getty Thesaurus, Alexandria Digital Library Gazetteer, NGA GNS, USGS, GNIS, World Gazetteer	Proposing a "Universal Database Architecture," a peer-to-peer system based on emerging internet data-sharing standards. Open Standards.	Unclear if this is currently used or just a proposal. "A Public Resource of Populated Places and Worldwide Administrative Divisions"
Falling Rain Gazetteer http://www.fallingrain.com/world/	Unknown.	Unknown.	Technical details unknown. Main page titled "Global Gazetteer Version 2.1."
Earth Search http://www.earthsearch.net/	CIA World Fact Book, NGA Earth-Info; USGS GeoNames.	Unknown.	Technical details unknown.
MultiMap World Map & Place-name Index http://www.multimap.com/index/	Unknown.	Unknown.	"Europe's most popular mapping website links to maps and location information."
Worldwide Index – Tageo.com http://www.tageo.com/index.htm	Geographic Area of Coverage of Tageo: Worldwide excluding Antarctica. This site provides a geographic coordinate database consisting :	Unknown.	"Tageo.com is a database of geographic coordinate information. Tageo.com provides information about 2,667,417 cities in the whole world."

Candidate Gazetteer / URL Link(s)	Data Sources	Standards	Comments
	- 5.4 million records from the NIMA GE0net Names Server - 2,7 million records of cities - 5000 records of administrative divisions - 188000 records of countries data - 7.9 gigabytes (7.9 Go) of land surface, shallow water, and shaded topography.		
Fuzzyg – The Fuzzy Gazetteer http://tomcat-dmaweb1.jrc.it/fuzzyg/query/	Unknown.	Unknown.	From the website: "The Fuzzy Gazetteer enables you to find geographic features even when you do not know their exact names. A list of similar names is returned, web-linked to the JRC Digital Map Archive of the European Commission. Searching 7,205,433 place names world-wide."
Country Specific Geographic Data Services			
EuroGeoNames (EGN) www.eurogeonames.org	Objective is to connect 5-10 national databases together.	OGC Standards.	EGN will establish a European Infrastructure of geographical names data and set up and interoperable Web (Gazetteer) service compliant to open standards. Open GIS Consortium (OGC) standards for web Feature Services. They are aiming to connect 5-10 national databases over the life of the project.
National Sultanate of Oman http://www.nsaom.org.om/english/services/	Unknown.	Unknown.	This agency is participating on a UN panel for geographic names.
Qatar - On line mapping service http://www.gisqatar.org.qa/MapServer/Default.htm	Unknown.	Unknown.	Qatar is a leader in GIS/Geospatial Technologies in the Mid-East. You can do internet based queries on their on-line map server - includes searching for place names.
Electronic Cultural Atlas Initiative - Iraq http://ecai.org/iraq/ Note see Electronic Cultural Atlas above.	An extension of the Alexandria Digital Library Project Gazetteer, a applied to art and culture.	Thesaurus.	Not really a gazetteer per se, but a collection of digital resources for preserving Iraq's cultural history. Historic place names are part of the collection.
Other Leads for Gazetteer Information			
United Nations Group of Experts on Geographical Names (UNGEGN) http://unstats.un.org/unsd/geoinfo/	The UNGEGN has a link to toponymic databases. http://www.zrc-sazu.si/ungegn/toponymi.htm	N/A	A working group UNGEGN is run by experts who want to improve international communication. Its organizational structure is that of a permanent UN Commission. Its objectives are geared to the improvement of global

Candidate Gazetteer / URL Link(s)	Data Sources	Standards	Comments
	This page contains links to toponymic data sets and digital gazetteers accessible on the World Wide Web. Based largely on the excellent list of links maintained by Peeter Päll at the site of the UNGEGN Working Group on Romanization Systems, the aim of this list is to target actual data sets.		communication through standardization of geographical names. Its Program is determined by its members, who convene regularly at sessions or represent their countries at 5-yearly standardization conferences. Activities are coordinated through the UNGEGN Secretariat at UN Headquarters in New York. Recently published, *Manual for the National Standardization of Geographical Names.* http://unstats.un.org/unsd/geoinfo/geog%20names%20final.pdf
Japan Geographical Survey Institute http://www.gsi.go.jp/ENGLISH/	Unknown.	Unknown.	Japan is participating in larger Gazetteer efforts in Asia.
Index of Place Names - University of North Carolina at Chapel Hill, College of Arts and Sciences Ancient World Mapping Center http://www.unc.edu/awmc/placenames.html	Unknown.	This is an index of place names referenced in a historical document collection.	Not really a gazetteer, but a source for historical place names, based on academic research. Focuses on Europe and the Middle East - Historical place names.
Google Directory of Place Names Links to many sites on place names. http://directory.google.com/Top/Science/Social_Sciences/Geography/Place_Names/	Unknown.	N/A	Google Directory search on “Place Names” provides a link to other websites that provide geographic place names.

APPENDIX D: BIBLIOGRAPHY

This bibliography was compiled as part of the first stage of the PATTON Alliance Gazetteer Evaluation Project. We compiled this from a variety of sources, including internet research and a bibliographic reference search. This bibliography was used to quickly understand the current state of research with regard to gazetteers.

Alexandria Digital Library (ADL) Gazetteer, 1999-. Santa Barbara CA: Map and Imagery Lab, Davidson Library, University of California, Santa Barbara. Copyright UC Regents. http://www.alexandria.ucsb.edu/gazetteer (ADL Gazetteer Development page with links to various clients and protocols that access the ADL Gazetteer, accessed 10/30/06).

Axelrod, Ammittai E., 2003. *On Building a High Performance Gazetteer Database.* Proceedings of the HLT-NACL2003 Workshop on Analysis of Geographic References, Alberta Canada, May 31, 2003. http://people.mokk.bme.hu/~kornai/NAACL/WS9/ws910.pdf (accessed 10/30/06).

Brauner, Daniela F., Marco A. Casanova, and Ruy L. Milidiu, 2006. *Towards Gazetteer Integration Through an Instance-based Thesauri Mapping Approach.* In Proceedings of the VIII Brazilian Symposium on GeoInformatics (GeoInfo 2006), Campos Do Jordao, Brazil. http://geoinfo.info/geoinfo2006/papers/p74.pdf.

Chaves, Marcirio Silveira, M. J. Silva, and and B. Martins, 2006. *A Geographic Knowledge Base for Semantic Web Applications.* In Proceedings of SBBD-05, the 20th Brazilian Symposium on Databases. : http://xldb.fc.ul.pt/data/Publications_attach/camera_ready_sbbd05.pdf (accessed 9/27/2006).

Clough, Paul. 2005. Extracting Metadata for Spatially-Aware Information Retrieval on the Internet. Workshop on Geographic Information Retrieval. In Proceedings of the 2005 Workshop on Geographic Information Retrieval, Bremen, Germany.

Duscaldi, Davide, Paolo Rosso, and Emilio Sanchis Arnal, 2006. *WordNet as a Geographical Information Resource*. Third International WordNet Conference (GWC-06) (downloaded 10/30/06 from http://nlp.fi.muni.cz/gwc2006/proc/index.html).

Fu, G., C. Jones, and A. Abdelmoty, 2005. *Bulding a Geographical Ontology for Intelligent Spatial Search on the Web.* In Proceedings of the IASTED International Conference on Databases and Applications, Innsbruck, Austria, February 14-16, 2005.

Goodchild, M.F., 1999. *The Future of the Gazetteer*. Presented at the Digital Gazetteer Information Exchange Workshop, Oct 13-14, 1999. Transcribed and edited from Audiotape. http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/Perspectives/brandt.htm.

Hill, Linda L., 2004. *Georeferencing in Digital Libraries* (guest editorial). In D-Lib Magazine, May 2004, 10 (5). http://www.dlib.org/dlib/may04/hill/05hill.html (All of the articles in this issue refer to the role of gazetteers in georeferencing applications.)

Hill, Linda L., 2000. *Core Elements of Digital Gazetteers: Place Names, Categories, and Footprints*. In J. Borbinha & T. Baker (Eds.), Research and Advanced Technology for Digital Libraries: Proceedings of the 4th European Conference, ECDL 2000, Lisbon, Portugal, September 18-20, 2000 (pp. 280-290). Berlin: Springer.

Hill, Linda L., 1999. *Foundations of Shareable Gazetteer Data*. Presented at the Digital Gazetteer Information Exchange Workshop, Oct 13-14, 1999. Transcribed and edited from audiotape. http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/Perspectives/hill.htm(accessed: 4/12/07).

Hill, Linda L. and Q. Zheng. 1999. *Indirect Geospatial Referencing Through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers.* In Proceedings of the American Society for Information Science Annual Meeting, Washington, D.C., Oct. 31- Nov. 4, 1999, pp. 57-69. http://www.alexandria.ucsb.edu/%7Elhill/paper_drafts/ASIS99_confpaper2_final.pdf (accessed: 4/12/07).

Hill, Linda L., J. Frew, and Q. Zheng, Q., 1999 *Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library*. In D-Lib (January 1999) (downloaded 8/22/2006 from http://www.dlib.org/dlib/january99/hill/01hill.htm l).

Hill, Linda L. 1999. *Gazetteer and Collection-level Metadata Developments*. In R. T. Kaser & V. C. Kaser (Eds.), Metadiversity. The Grand Challenge for Biodiversity Information Management through Metadata. The Call for Community. In Proceedings of the Symposium sponsored by the U.S. Geological Survey Biological Resources Division. and the National Federation of Abstracting and Information Services (pp. 141-145): NFAIS. http://www.nfais.org/publications/metadiversity_preprints23.htm.

Hill, Linda L. and Q. Zheng, 1999. *Indirect Geospatial Referencing Through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers*. In Proceedings of the American Society for Information Science Annual Meeting, Washington, D.C., Oct. 31- Nov. 4, 1999, pp. 57-69. http://www.alexandria.ucsb.edu/%7Elhill/paper_drafts/ASIS99_confpaper2_final.pdf.

Irie, R.and B. Sundheim. 2004. *Resources for Place Name Analysis*. In Proceedings of the 2004 International Conference on Language Resources and Evaluation - LREC2004, pp. 317–320.

Jones, C.B., 2002. Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT Project. In SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 11-15, 2002, Tampere, Finland, ACM Press, pp.387-388 (downloaded 11/6/2006 from: http://www.geo-spirit.org/publications/SPIRIT_project_overview.pdf).

Kornai, Andras and Beth Sundheim, 2003. *Workshop on the Analysis of Geographic References*, May 31, 2003. In NAACL-HLT 2003, Edmonton, Alberta, Canada. http://people.mokk.bme.hu/~kornai/NAACL/WS9/paper.html. Conference website: http://www.kornai.com/NAACL (accessed 9/27/2006).

Landcaster , Lewis, 2002. *A Multilingual Gazetteer System for Integrating Spatial and Cultural Resources*. Electronic Cultural Atlas Initiative, UC Berkley. http://ecai.org/projects/gazetteer/nsf_multisys_abstract.html (accessed 9/27/2006).

Linton, Frank, 2005. *Accelerating GEOINT Analysis: User Feedback to Improve Geospatial Place-Name Services*. MITRE Corporation White Paper. http://www.mitre.org/work/tech_papers/tech_papers_05/05_0247 (accessed 9/302006).

Lucas, Peter, M. Balasubramanya, D. Widdows, and M. Higgins. *The Information Commons Gazetteer: A Public Resource of Populated Places and Worldwide Administrative Divisions*. 2006 Language Resources and Evaluation Conference (LREC), May 24-27, 2006, Genoa, Italy. www.maya.com/web/what/papers/maya_infocommons_gaz.pdf (downloaded 9/25/2006).

Mardis, Scott and John Burger, 2005. *Design for an Integrated Gazetteer Database: Technical Description and User Guide for a Gazetteer to Support Natural Language Processing Applications*. MITRE Corporation White Paper. http://www.mitre.org/work/tech_papers/tech_papers_06/06_0375/ (accessed 9/30/2006).

MetaCarta, Inc., 2005. *A White Paper on MetaCarta's Technology and Products.* http://www.metacarta.com/docs/Corporate_White_Paper.pdf (accessed 10/30/2006).

MetaCarta, Inc., 2005. *MetaCarta's Technology and Products: A Public Sector White Paper.* http://www.metacarta.com/docs/Public_Sector_White_Paper.pdf (accessed 10/30/2006).

Rauch, Erik , Michael Bukatin, and Kenneth Baker, 2003. *A Confidence-based Framework for Disambiguating Geographic Terms.* Paper presented at the May 31, 2003, Workshop on the Analysis of Geographic References at the NACC-HLT 2003 Human Language Technology Conference. May 27-June ?, 2003, Edmonton, Canada. http://www2.sims.berkeley.edu/research/conferences/hlt-naacl03/. Paper: http://people.mokk.bme.hu/~kornai/NAACL/WS9/ws917.pdf. Accessed: 4/13/07.

Rugg, Bob, 1999. *Emerging Geographic Information Standards and Gazetteer Data*. Presented at the Digital Gazetteer Information Exchange Workshop, Oct 13-14, 1999. Transcribed and edited from audiotape. http://www.alexandria.ucsb.edu/~lhill/dgie/DGIE_website/Perspectives/rugg.htm (accessed 9/30/2006).Smith, David A., 2002. *Mining Gazetteer Data From Digital Library Collections*. In proceedings of the Networked Knowledge Organization Systems (NKOS) Workshop, July 18, 2002, Portland, Oregon. Joint Conference on Digital Libraries (JCDL 2002) on Digital Gazetteers. http://nkos.slis.kent.edu/ (accessed 10/26/2006).

Souza, L. A., C. A. Davis Jr., K. A. V. Borges, T. M. Delboni, and A. H. F. Laender, 2005. *The Role of Gazetteers in Geographic Knowledge Discovery on the Web,* In Proceedings of the 3rd Latin American Web Congress, October 31-November 2, 2005, Buenos Aires, Argentina. http://ieeexplore.ieee.org/iel5/10609/33518/01592372.pdf?tp=&arnumber=1592372&isnumber=33518 (downloaded 11/6/06).

Sundheim, Beth M., Scott Mardis, and John Berger. 2006 *Gazetteer Linkage to Word Net*. Third International WordNet Conference (GWC-06). http://nlp.fi.muni.cz/gwc2006/proc/ (downloaded 10/30/06).

Toral, Antonio, and Rafael Munoz, 2006. *A Proposal to Automatically Build and Maintain Gazetteers for Named Entity Recognition by Using Wikipedia*. Workshop on New Text, 11th Conference of the European Chapter of the Association for Comuputational Linguistics. Trento, Italy, April 2006. http://www.sics.se/jussi/newtext/working_notes/09_toral_munoz.pdf (accessed 11/6/06).

Uryupina, Olga, 2003. *Semi-supervised Learning of Geographic Gazetteers from the Internet*. In Proceedings of the HLT-NAACL Workshop on the Analysis of Geographic References, Edmonton, 2003.

Ourioupina, Olga, 2002.. *Extracting Geographical Knowledge from the Internet*. In Proceedings of the ICDM-AM International Workshop on Active Mining, Maebashi, 2002.

DISTRIBUTION

1	MS 1221	Bill Cook, 05630
1	MS1235	Judy Moore, 05634
1	MS1235	Rolf Carlson, 05634
1	MS0763	Steve Ratheal, 06453
4	MS0763	Denise Bleakly, 06453
2	MS9018	Central Technical Files, 8944
2	MS0899	Technical Library, 9536